Peer Evaluation of Teachers in Maricopa County's Teacher Incentive Fund Program

Anthony Milanowski, Westat
Herbert G. Heneman, III, University of Wisconsin-Madison
Matthew Finster, Westat

Introduction

This case study describes the peer evaluation system the Maricopa County Educational Services Agency (MCESA) is using in the districts participating in its TIF 3 and TIF 4 grants. This brief first discusses the potential advantages and challenges of implementing peer evaluation and then describes the design and evolution of MCESA's peer evaluation process and what is currently known about its impact on teachers and school leaders. It concludes by presenting the lessons MCESA learned in operating the program since the 2012–13 school year.[1] MCESA has been particularly thoughtful in designing a peer evaluation approach that fits the needs of its districts. It has successfully addressed most of the challenges of implementing peer evaluation. While the exact design of this program will not be applicable everywhere, MCESA has learned several lessons that could help other grantees design and implement peer evaluation programs that would further their grant's objectives.[2]

Peer evaluation involves colleagues observing and rating teaching practice and using these ratings as an input in determining the final evaluation rating. For example, if a district requires three observations of a teacher's practice, the peer may do one or two, and the teacher's supervisor (typically a principal or assistant principal) would do the others. Then the supervisor would use the results of each observation to determine the practice rating, or the ratings of each observation would be combined (for example by averaging) to produce a final practice rating.

Peer evaluation can be part of a peer assistance and review (PAR) program, which typically focuses on new or struggling teachers and gives the peer evaluator the responsibility to recommend whether the teacher be continued or dismissed. The final decision is often made by made by a joint association-management committee that also governs and monitors the program (Project on the Next Generation of Teachers, n.d.). However, as in Maricopa County, peer evaluation can be used without being part of a PAR process.

---

[1] The description of Maricopa County's TIF peer program is based on documents provided by program administrators and conversations with the TIF director, TIF 3 and 4 program managers, peer evaluators, field specialists, the evaluator of the peer evaluators, and two principals. We summarized this information and shared an initial draft with staff of the Maricopa County Educational Services Agency, who suggested corrections where appropriate.

[2] Note that forms of peer evaluation have also been used by other Teacher Incentive Fund grantees, including the Austin (Texas) Independent School District, the District of Columbia Public Schools, Hillsborough Public Schools in Florida, and the Pittsburgh (PA) Public Schools.

Advantages and Challenges of Peer Evaluation

Peer evaluation has several potential advantages for use in TIF-supported teacher evaluation systems:

- Peer evaluators can share the burden of multiple observations with school administrators. Multiple observations support continuous improvement because teachers can apply the feedback they get and to try out the suggestions, then get more feedback on how they are doing and how to further refine their teaching, as well as providing a more reliable estimate of teaching performance.

- Peer evaluators, if matched with teachers on grade level and subject experience, could have a deeper understanding of the teachers' subjects and grades than school administrators. In particular, peer evaluators might have more insight into the content knowledge and content-specific pedagogy required. Peer evaluators could be both more accurate observers of these aspects of teacher practice and could provide more credible feedback as well.

- Peer evaluators could encourage school administrators to make more accurate ratings. Sharing the responsibility for evaluation could make it easier for school administrators to honestly rate low-performing teachers as well as discouraging them from acting on any biases they have for or against particular teachers.

While peer evaluation seems promising, there are several challenges to meet in order to make it work.

- Teachers may be reluctant to have peers in their classroom and subject themselves to potential peer disapproval. They may believe that a peer evaluator from outside the school will not understand the school and the challenges its teachers face. Veteran teachers may resent younger peer evaluators judging their teaching or even providing suggestions. Less skilled teachers might fear that peer evaluators will be better able to detect problems in their teaching.

- School leaders could perceive that sharing evaluation responsibilities with peer evaluators could diminish their authority to run their schools. They could be concerned that evaluators from outside the school would apply a stricter or more lenient standard.

- Teacher organizations could also be concerned that peer evaluation would reduce group solidarity and blur the line between teachers and management. If a peer evaluator gives a low rating that affects another teacher's job, the organization might have to choose whom to support if the rating was disputed.

- Peer evaluators could have difficulty balancing evaluative and coaching roles. Teachers may expect that a peer's primary role is to help them, and many peers may be more comfortable providing help and formative feedback than an evaluation rating that could have consequences for fellow teachers. Yet much of the benefit of peer evaluation could be lost if the evaluative role is removed, reducing the motivation of the teacher to take the feedback and coaching seriously. Writers on performance evaluation have long recognized potential conflict between the evaluator and coach roles (e.g., Meyer, Kay, & French, 1965; Milanowski, 2005; Popham, 1988).

- Ensuring peer evaluators and supervisors rate consistently and provide consistent feedback is a challenge. Peer evaluation will lack credibility and confuse teachers if the two types of evaluators apply different standards when rating or recommend different changes in practice.

- Peer evaluation can be expensive. Not only do evaluator's salaries need to be paid, but there are also training costs.

This report describes how Maricopa County addressed these challenges.

Peer Evaluation Initial Design and Implementation in Maricopa County

Maricopa County's TIF Districts

Maricopa County's TIF districts cover much of the Phoenix, Arizona, metropolitan area. Including both TIF 3 and 4 cohorts, 12 districts with a total of 82 schools are participating in the TIF project. The districts range in size from 1 to 19 schools, all of which had high proportions of students in poverty. Overall, approximately 2,200 teachers participated in the peer evaluation process.

Program Design

The Maricopa County TIF program designed peer evaluation into its teacher evaluation process from the beginning. The initial impetus for peer evaluation came from the insight that in order for teachers to improve their instructional practice, they need content-specific feedback from someone with a high level of skill both in teaching and in sharing actionable feedback with teachers. Other stakeholders in the planning process, including teacher and administrator organizations, the state education agency, and representatives from the districts interested in joining the TIF application, also had concerns. Local teacher organization members raised the issue of the quality of feedback school administrators could provide and put forward the concept of peer support for teachers undergoing evaluation. Teachers were also concerned that one or two observations would not provide enough interaction with the evaluator to improve performance. Stakeholders recognized that, across the multiple districts that would be participating in the TIF 3 grant, school administrators varied substantially in their knowledge of the content taught by the teachers in their schools, as well as their ability to provide useful feedback. Too often, teachers did not see the feedback many administrators provided as useful. During the planning process for the TIF 3 grant, the stakeholders agreed that peer evaluation could address these concerns.

To provide useful, content-specific feedback, MCESA staff came up with the idea of recruiting a cadre of "rock star" teachers. As well as providing better feedback, this cadre could also help improve the accuracy of performance ratings, based on their specific knowledge of the content area of the teachers they would observe. Peer evaluation could also be used to reduce time demands on administrators made by the initial decision to require five observations. (Five were initially chosen to provide enough observations for teachers to apply the feedback they got and have the evaluator observe any improvement and went beyond a change in state law requiring two observations.) To keep principals in the loop, the design committee divided each teacher's observations between the principal and the peer evaluator. Another initial decision was to have the peer evaluators' ratings count in determining teachers' summative evaluation ratings, as well providing formative feedback. Since most teachers have some room for improving their practice, all teachers would receive peer evaluations, not just new or struggling teachers. The peer evaluators would devote their full time to teacher evaluation, without having classroom teaching responsibilities.

Initial Implementation

Peer evaluation began in 2012–13 in the five districts participating in the TIF 3 grant. Both building administrators and peer evaluators used the TIF observation rubric MCESA and its district partners developed. The new rubric covers 22 components of teaching across six domains and has five performance levels. The teaching practice rubric, called the Learning Observation Instrument, is available at: **http://mcesa.schoolwires.net//site/Default.aspx?PageID=316**.

The observation process begins with a preconference at least 24 hours before each observation, followed by the observation and a postconference. Together these activities make up an observation cycle. The preconference collects some evidence relevant to the evaluation standards, and helps the evaluator understand the lesson to be observed. After the observation, the peer evaluator conducts a postconference, in which he/she shares the ratings from the

observation. The evaluator and teacher review the evidence gathered at the preconference and observation. At the postconference, the teacher also shares an analysis of the lesson and any documentation relevant to the lesson. The evaluator also identifies and discusses a reinforcement (an observed action taken by the teacher that would likely have a positive influence on student learning) and a refinement (a suggestion for an action that, if implemented by the teacher, would have a positive impact on student learning).

Peer evaluators do two to three of the required five observations of each teacher, with the building administrator doing the rest. After each observation cycle, peer evaluators share their ratings with building administrators as well as with the teacher they observe. Peer evaluators also enter their ratings into an evaluation information management system. Building administrators do not, however, have to share their ratings with peer evaluators. Peer evaluators' ratings count the same as building administrators' ratings in determining a teacher's instructional practice rating. The ratings made for each observation on each element are combined across observations to determine each element's final score.

Initial Implementation Challenges and Changes in Program Design
Despite careful planning, several problems arose during the first months of peer evaluation in 2012 in the TIF 3 districts:

- Initially, peer evaluators were often rating teachers lower than building administrators and lower than these teachers had been rated in the past. This upset teachers, building administrators, and even some school boards. Many building administrators felt they had to protect their teachers, especially when initial evaluations by peer evaluators were lower than their own ratings and lower than their teachers had received in the past.

- In order to schedule the peer evaluators most efficiently, many teachers were observed by more than one peer evaluator. This limited the peer evaluators' opportunity to build a trusting relationship with the teachers.

- One consequence of the peer evaluators' content specialization was that in some schools several different peer evaluators were required to cover all of grade-level/subject combinations of teachers in the schools. Building administrators were surprised that so many different peer evaluators were coming to their buildings, which made many uneasy. It was hard for them to establish a relationship with the peer evaluators, which limited their level of trust in the peer evaluators' ratings.

- Despite a rigorous selection and training process, some peer evaluators had performance problems, including difficulty interpreting evidence and following the rubrics, interacting with teachers, and gaining building administrator trust.

MCESA and its partner districts responded quickly to these problems.
MCESA made its peer evaluator training more rigorous (as described below), and peer evaluators received more training on establishing relationships with teachers and building administrators. Field specialists did co-observations with those having difficulties. Some peer evaluators were replaced. Peer evaluators were allowed to spend more time coaching teachers outside of the evaluation cycle. This gave teachers concerned about their ratings more guidance on how to improve their performance. It was also welcomed by the peer evaluators, who wanted to provide more coaching. Peer evaluators began to take more active roles in professional development planning and working with school professional learning communities. A related change was to have peer evaluators design professional development modules that they could deliver along with the building administrator. These changes helped to strengthen relationships with building administrators.

MCESA also recognized that it needed much more communication with teachers, building administrators and school boards to ensure support for peer evaluation. To communicate with school boards, MCESA's TIF field specialists (the grant program's primary contact person with each district) attended board meetings and made presentations on TIF and the peer evaluation program. Field specialists and peer evaluators also were encouraged to communicate about the program directly to building administrators, instead of relying on communication with district central office staff to filter down. MCESA asked peer evaluators to meet with teachers before the first observation cycle, to introduce themselves and explain the process, in a pre-preconference, and to meet and introduce themselves to building administrators.

A dispute resolution process was developed for teachers or building administrators to raise disagreements about evaluation ratings made by peer evaluators. Teachers or building administrators first request a conference with the peer evaluator, who discusses the ratings given and explains how the evidence gathered justifies them. However, peer evaluators are not allowed to change their ratings, so that they can't be pressured to change scores. If this conference does not resolve the issue, the MCESA field specialist becomes involved. Almost all disputes have been resolved at this point if not at the initial conference. The care with which peer evaluators record evidence and the option of having the peer evaluator provide additional professional development to the teacher contributed to resolving these disagreements.

Another change was to reduce the number of observation cycles to four in some districts. This allowed caseloads to be reduced and provided peer evaluators with more time for coaching and professional development at the schools. In one of these districts, the peer evaluator and building administrator began doing each of the four observations together. They discuss their ratings, but do not have to come to agreement. Co-observation provides a way for the peer evaluators to educate building administrators about how to observe and use the rubric and for building administrators to educate the peer evaluators about the culture of the school.

MCESA decided to change scheduling to reduce the number of different peer evaluators coming and going to and from each school and to decrease the number of teachers observed by more than one peer evaluator. This helped to ensure teachers received consistent feedback and to allow a relationship of trust to develop. It also helped peer evaluators and building administrators to become more familiar with each other. The cost of these changes was less balanced peer evaluator caseloads. For the 2013–14 school year, caseloads varied from 30 to 75, though 50 teachers per peer evaluator is considered the ideal. Peer evaluators with lower caseloads had opportunities to work on other projects, such as developing professional development programs related to the vision of teaching behind the Learning Observation Instrument.

Another early implementation issue was whether teachers placed on a performance improvement plan by their district should have a peer evaluator. These teachers were typically being considered for dismissal, and some districts did not want to change the expectations of the process for teachers or evaluators. Districts were also concerned that teachers might get different feedback from peer and administrator evaluators. In response to these concerns, some districts do not use peer evaluation in these cases.

MCESA's program managers were able to react quickly because they had established multiple feedback loops to keep informed about how peer evaluation was being implemented. Field specialists provided feedback about how district administrators, building administrators, teachers, and peer evaluators viewed the program. Biweekly peer evaluator meetings offered program managers a regular opportunity to hear from peer evaluators. MCESA also conducted surveys of teachers and principals about what they experienced with peer evaluation and provided the results to the peer evaluators as well as program managers.

MCESA implemented these changes in 2013–14, the second year for TIF 3 districts and the first year for the TIF 4 grantees. Along with more experience with interacting with peer evaluators, the changes substantially reduced the initial opposition from teachers and building administrators.

Though peer evaluation ran much more smoothly in the 2013–14 school year, and with much less opposition, some challenges remain:

- Some senior teachers still resent younger peer evaluators coming in to evaluate them.

- Some schools are still not welcoming to peer evaluators, even for providing professional development.

- Mid-year changes in the peer evaluator assigned to a teacher are still sometimes needed to cover all of the observation cycles or in cases of peer evaluator turnover.

- There is a large amount of work involved to match peer evaluators to teachers and schedule observations.

- MCESA is still working on how to balance the caseloads of the peer evaluators. Many peer evaluators have substantially larger or smaller caseloads than the ideal of 50, and some schools still have as many as 12 peer evaluators involved in the evaluation process. It has proven difficult to balance the need for a content match between the teacher and the peer evaluator with the need to minimize the number of evaluators visiting each school.

- In some schools, peer evaluation has threatened to bring performance issues with teachers that building administrators have not addressed into the open. Some administrators believe they will have trouble replacing poorer teachers pushed out by more rigorous evaluation due to competition with more wealthy districts and those with less rigorous evaluation systems.

Human Capital Management for Peer Evaluators

From the beginning, MCESA and its partner districts recognized that the selection, training, and evaluation of peer evaluators would be an important influence on the success of the initiative. They made substantial investments in developing these human capital management processes specifically for peer evaluators.

Peer Evaluator Recruitment and Selection
Recruiting a sufficient number of qualified candidates was initially challenging. Though MCESA had about 300 applicants, many did not have the combination of content expertise and skill in providing content-related feedback, goal setting, and coaching. It proved to be difficult to find enough recruits in some content areas. MCESA found itself in competition with member districts that employed coaches, and with another TIF grantee in the area that was also looking for coaches. MCESA found that it had to do a lot of networking to find qualified candidates, using staff's connections and word of mouth. The recruitment message emphasized that working as a peer evaluator is an experience candidates can't get elsewhere and can take to other districts.

MCESA uses a multi-step selection process to assess the degree to which candidates for peer evaluator positions possess the competencies needed. Besides the standard job interview, MCESA asked candidates to:

- Complete written exercises describing their qualifications, and how they would create professional development in their content area based on one of the state student content area standards

- View a video of teaching and describe how they would coach the teacher depicted in the video

- Conduct a simulated postobservation conference, with the role of the evaluated teacher being played by another peer evaluator

- Participate in a group activity focused on developing a professional development opportunity based on simulated disaggregated performance evaluation data

- Meet with district staff and begin to build a relationship with them.

Multiple evaluators scored all the activities using rubrics.
Both MCESA staff and the peer evaluators themselves consider this a rigorous selection process that also introduces candidates to the demands of the peer evaluator role. Most of the peer evaluators hired to date have had prior leadership experience, as an instructional coach, professional development developer, or even as a building administrator.

Peer Evaluator Training
Training begins with 30 hours of calibration training involving observing and rating performance as depicted in videos of teachers. Peer evaluators also receive training on conferencing with teachers, including setting goals based on observation data, providing honest feedback, coaching, and building relationships with both the teachers they evaluate and the building administrators they work with. Feedback and coaching training includes role playing to provide additional realism. The training also includes co-observing in live classroom settings, with emphasis on scripting, scoring, selection of postconference objectives, and questioning strategies. Peer evaluators also received training on justifying their ratings based on the evidence they collect.

Training is ongoing. Every other Friday during the school year, peer evaluators gather for a full day of training, including rating additional videos and discussions of how to interpret the evaluation rubric (e.g., how it applies to specific student populations, such as special education students). In order to secure enough videos to use for training, teachers in participating districts were offered feedback and professional development in return for sharing a video. This enabled the trainers to assemble a set of videos representing a wide range of performance levels. Even so, trainers had to stage some videos to secure enough examples of top-level teaching.

Peer evaluators also get informal training through their everyday interactions with their colleagues. MCESA decided to base all of the peer evaluators in a single location in its headquarters building, to allow them to discuss issues and consult with each other informally. One peer evaluator described working with fellow peer evaluators as the equivalent to a university course in teaching improvement.

Peer Evaluator Evaluation
Early on, MCESA and its partner districts recognized the need for a performance evaluation process tailored to the peer evaluator role. A specific evaluation rubric was developed for observing peer evaluators' practice (see the overview in Table 1 below.)

Table 1. Overview of Peer Evaluator Observation Instrument

Domain	Elements	Evidence Collection During
Pre- & Postconference: Data Gathering	Focus on conference objectives	Observation of pre- & postconferences Observation
	Engages teacher in reflection	
	Use of questions	
Pre & Postconference: Reinforcement and Refinement	Provision of feedback	
	Sharing of improvement strategies	
Conference Process	Conference is clear, well, paced, relevant, & engaging	
	Use of oral and body language	
Mutual Trust & Respect	Active listening & establishing a positive relationship	
Observation & Evaluation of Instruction	Scripting accuracy Rating on annual assessor certification	Review of artifacts Annual certification assessment

The elements for peer reviewers that cover observing and evaluating teachers are similar to those in the rubric for school leaders. This sends the message that both types of evaluators are held accountable in the same way for high-quality observations and accurate ratings of teachers.

One specialized evaluator observes all the peer evaluators and rates their practice using the rubric. She observes two sets of pre- and postconferences for each peer evaluator and reviews the artifacts the peer evaluator collects. She reviews two sets of scripts and artifacts to see how well ratings are justified with the evidence collected. Though watching peer evaluators observe was considered, this would be logistically difficult with 40 peer evaluators. Peer evaluators' participation in frequent calibration sessions and a yearly calibration test are considered as measures of rating accuracy. The test includes scripting and rating two videos and suggesting feedback and conferencing topics. Like the teachers they evaluate, the peer evaluators receive a practice score based on their ratings on each of the standards.

Peer evaluators also receive evaluations based on student achievement growth. For 2012–13, their student growth scores were based on the schoolwide growth measures of the schools in which the teachers they evaluated worked. Each of the two schoolwide growth measures for each school was weighed by the percentage of observations the peer evaluator did in each school. Each weighted measure was then summed across schools and the two are combined with the practice rating. The practice rating was weighted at 75% and the two school growth measures were weighted at 15% and 10%. The resulting score was compared to a table of score ranges that gives the overall summative rating associated with each point score. For 2013–14, the student growth attributed to the teachers peer evaluators worked with was used for the student growth component rather than schoolwide growth. The growth score for each teacher was weighted by the time the peer evaluator spends with each. Student growth was weighted 25%, practice 75%.

Using student growth to evaluate peer evaluators has been somewhat controversial. Peer evaluators have been concerned that they have limited influence on the growth of the students of the teachers they evaluate. However, MCESA and its partners believe that if they hold teachers accountable for student growth, the peer evaluators should be held accountable as well. They also believe that peer evaluators' credibility with the teachers would also be reduced if the peers were not evaluated based in part on growth. Peer evaluators recognized this, though it has not eliminated their concern.

Peer Evaluator Compensation
As mentioned above, peer evaluators are MCESA employees. They are not paid on a schedule like teachers in the districts. They are hired at a salary depending on qualifications and prior salary history that is within a range of approximately $60,000 to $81,000 per year. (For reference, the average teacher salary in Arizona was $48,885 for 2012–13.) Peer evaluators are eligible for performance-based compensation based on their overall effectiveness ratings that is comparable to that available to teachers.

Initial Effects of Peer Evaluation

While MCESA does not yet have systematic evidence of the effects of peer evaluation on teacher performance, peer evaluators, building administrators, and field specialists all report some early positive impacts on both building administrators and teachers.

With respect to teachers, anecdotal evidence suggests that:

- Peer evaluators have been able to provide content-specific professional development in districts or schools where this had been rare in the past, due to limited school or district resources.

- Feedback from a credible content expert encourages teachers to seek out assistance from teaching peers and coaches at their schools.

- Feedback and coaching have helped teachers improve specific aspects of their instruction. Peer evaluators and building administrators recognized improvements in practice between the initial observation cycle and later cycles.

- Feedback and coaching from peer evaluators, coupled with the observation rubric, provides a roadmap for new teachers to get up to speed.

- Peer evaluation, in conjunction with the more rigorous evaluation rubric, has encouraged some low-performing teachers, and teachers who do not want to change their practice, to leave the TIF districts.

- In some cases, the peer evaluator's additional evidence and support provided made it easier to dismiss teachers for poor instruction.

Of course, peer evaluation does not always lead to practice improvements. Peer evaluators acknowledge that some teachers remain resistant to feedback, and some building-level administrators are hesitant about peer evaluators providing teacher professional development in their schools.

Because Maricopa County's TIF districts implemented a more rigorous evaluation rubric as well as peer evaluation, many of the effects on teachers cannot be attributed to peer evaluation alone. But it does appear that peer evaluation has facilitated the implementation of a more rigorous evaluation process. Not only do peer evaluators allow a greater number of observations to be done, they may influence building administrators to be more rigorous. Peer evaluators also provide the coaching and content-specific professional development that gives teachers the skills to improve. Having the resources available to improve performance where needed has also likely increased teachers' acceptance of the more rigorous evaluation process.

For building administrators, positive impacts include:

- Allowing the building administrators to complete the required number of observations without overburdening them. Principals have had additional time to get into classrooms and do informal observations, walk-throughs, and coaching.

- Peer evaluators have helped school administrators improve school-based professional development, especially in smaller schools and districts, by helping to plan and conduct professional development at schools as well as providing content-specific professional development to individual teachers.

- Co-observation with peer evaluators and interacting with them around the evidence for ratings has encouraged building administrators to improve their feedback and instructional leadership skills.

- Conversations with the peer evaluators around evaluation ratings and the supporting evidence helped building administrators understand the rubric. Better understanding should lead to more consistency among raters and potentially higher inter-rater agreement. Initial analyses by the grantee's evaluator showed that the average percentage of exact agreement in 2013-–4 was 73 percent, and agreement within one level on the five-level Learning Observation Instrument rating scale was 97 percent.

- Peer evaluators provide confirmation of building administrators' observations by a content expert, which increases their confidence in their ability to use the rubric and identify more and less effective teaching.

- Peer evaluation has made it easier for building administrators to work with low-performing teachers by providing someone with content expertise and from outside the building to confirm their observations. This sends the message to low-performing teachers that the reason for a low rating is low performance, not the principal's bias against the teacher.

Peer evaluation, coupled with evaluator training and holding building administrators accountable, could also have discouraged building administrators' natural tendency toward rating leniency. Figure 1 below shows the distribution of practice ratings from the TIF 3 districts for the 2012–13 school year.

Figure1: Distribution of Teacher Practice Rating in Maricopa County TIF Districts, 2013–14 School Year
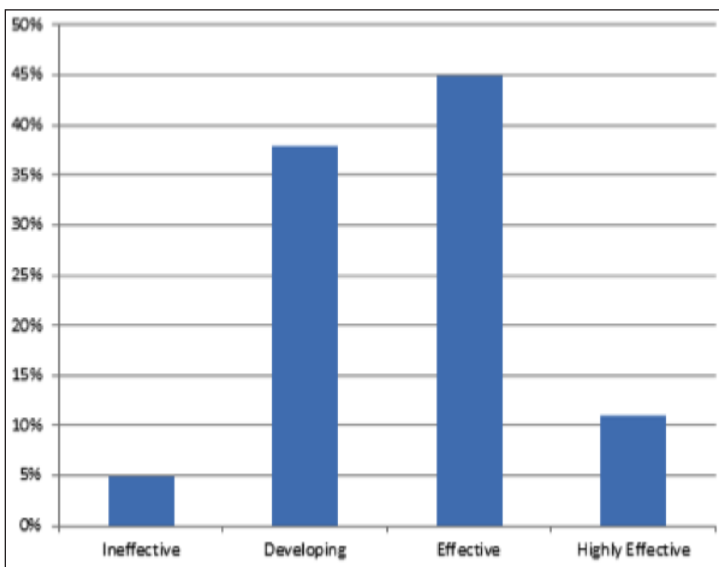
This distribution appears to be less weighted to the high end of the rating scale than for many other districts, and more nearly approximates the more symmetric distribution of teacher effectiveness many have expected more rigorous evaluation systems to yield.

One unexpected consequence of the more rigorous evaluation process is that it can encourage good teachers to leave. As one peer evaluator observed, "Some high performers leave, too. They don't want to jump through the hoops." Teachers who receive high evaluations are more marketable outside the TIF districts because the Maricopa evaluation system is perceived as rigorous, and Arizona law allows for prospective employers to receive teachers' past evaluations. High ratings give a teacher an advantage in competing for outside positions. Teachers who have done well

are also aware that they are marketable, and some leave the districts for jobs in districts with less demanding student populations. Some of these effects could be counteracted when the TIF districts fully implement salary systems that give greater weight to performance when determining salary increases.

MCESA and its district partners believe that more rigorous evaluation that includes the use of peers is a means to improve student achievement rather than an end in itself. As one field specialist put it. "We ask building administrators to support change in the evaluation process but we don't know yet if it works. Will teachers rated at high levels in the rubric get higher student achievement? They are taking it on faith now. They want to see that the evaluation results are correlated with student achievement growth." MCESA is working with its TIF program evaluator to examine this question and more systematically assess the effects of the evaluation process.

Costs and Sustainability

Maricopa County developed and refined its peer evaluation program over two years and used it again in the 2014–15 school year. Participating districts and other stakeholders appear to value peer evaluation, and MCESA is likely to sustain it through the end of the TIF 4 grant. Two issues will have to be addressed to sustain the program beyond TIF: cost and control.

Even with the relatively large caseload of each peer evaluator, peer evaluation costs are substantial. Doing two to three observation cycles for the approximately 2,200 teachers in the participating districts requires 40 peer evaluators. They are supported by one full-time position that evaluates them and oversees ongoing training. Based on information received from MCESA, we estimate the overall cost of the program at $4.0 million, or approximately $1,800 per year per teacher evaluated. When the TIF grants end, MCESA and/or the participating districts will have to absorb these costs. The current plan is to make the peer evaluators available to districts in return for a fee to participate in the program. MCESA believes enough districts will participate to sustain at least some version of peer evaluation.

Whether districts decide to participate could also depend on how they see the trade-off between the advantages of taking control of peer evaluation themselves versus enjoying the economies of scale MCESA's running of the program provides. According to MCESA staff, some districts have been interested in housing the program locally to customize the program to their needs and context. However, districts, especially the smaller ones, recognize that it would be difficult to support content specialists in all areas and to provide the level of training and support to peer evaluators that a centrally administered program can.

Whether districts participate could also be influenced by how well peer evaluation coupled with the more rigorous evaluation process actually improves instructional practice. Though there is little if any research that compares the effectiveness of peer evaluation to other forms of professional development, it could be more cost effective. Some studies (e.g., Chambers et al., 2008; Miles et al., 2004; Odden et al., 2002) have suggested that district budgets have substantial amounts dedicated to many different uncoordinated professional development efforts, and districts could reallocate some of these funds to support peer evaluation. Reallocation of ESEA Title II funds is also a possible source of support for peer evaluation.

Lessons From Maricopa County's Peer Evaluation Experience

MCESA staff identified several lessons they have learned about peer evaluation:

• Content knowledge is essential to credibility with teachers and building administrators.

• Credibility and support also depend on building a relationship of trust and familiarity with the teachers and building administrators. Intensive efforts are needed to establish these relationships. Table 2 shows the most important strategies the Maricopa County has used.

Table 2: Communication and Relationship-Building Strategies Maricopa County TIF Used to Support Peer Evaluation

Peer Evaluators	• Hold a conference with teachers before the first observation cycle to introduce themselves and answer teacher questions about the evaluation process • Meet with building administrators before the first observation cycle to introduce themselves and answer questions about the evaluation process • Participate in school walk throughs and attend school staff meetings • Conduct co-observations with building administrators • Co-present professional development modules with building administrators • Attend school professional learning community meetings
Field Specialists and MCESA Program Administrators	• Meet with building administrators to explain peer evaluation and are the first responders to discuss and resolve problems (field specialists) • Provide training on the evaluation rubric and do co-observations with building administrators (field specialists) • Attend school board meeting to explain evaluation program and TIF in general • Access a brochure on the TIF website that summarizes the rational for and operation of peer evaluation • Develop web profiles and fact sheets about the individual peer evaluators to help teachers and building administrators learn more about them • Match and schedule peer evaluators to limit the number of peer evaluators coming into buildings and maintain a single peer evaluator for each teacher • Include video testimonials from building administrators and teachers on the TIF website

• It is important to hire the right people. In addition to having content and coaching skills, peer evaluators have to be able to establish relationships, adapt to program change, feel comfortable giving low scores when needed, and stand up to challenges.

• Peer evaluators need to receive training to build relationships as well as to become experts in the evaluation rubric and have high levels of inter-rater agreement.

• Having a central home base for peer evaluators allows them to interact with one another, fosters mutual support, and helps them learn from each other. MCESA staff found that having a dedicated office or space for peer evaluators in each school also helps establish relationships with school staff.

• Aligning the evaluation of teachers, building administrators, and the peer evaluators reinforces each. Each group's observation rubrics have similar structure and use consistent terminology. The rubrics for peer evaluators and school leaders share similar language about observation and feedback, and school leaders are evaluated on how well they evaluate teachers, including how well they use feedback from peer evaluators to coach and plan professional development for teachers. The peer evaluation results are also used in teachers' Educator Goal Plan, a professional development plan that requires the teacher and the building administrator to establish one goal based on rubric ratings (as well as a student achievement goal).

Meeting the Challenges of Peer Evaluation

The design and implementation of peer evaluation in Maricopa County's TIF districts also has addressed the three challenges mentioned in the introduction: gaining and maintaining support from stakeholders, finding the right balance between the peer evaluators' evaluative and coaching roles, and maintaining consistency between peer and supervisor evaluators.

Stakeholder Support

Two features of the Maricopa program reduced if not eliminated concerns from teacher organizations. First, peer evaluators are employees of MCESA, not the districts. Teacher organizations do not represent the peer evaluators, so that any negative judgments they make about teachers do not pit one member against another. Second, the program has put substantial emphasis on providing teachers with resources to improve performance, so that peer evaluation is not just about judging teachers. The shift toward more coaching and targeted professional development from peer evaluators likely helped maintain teacher organization support.[3]

Potential opposition from administrators was likely mitigated by emphasizing the content expertise of the peer evaluators. Field specialists told us that many building administrators recognized that they did not have the content expertise to provide feedback and coaching to all teachers on all aspects of the observation rubric. These administrators did not regard peer evaluators as threatening their self-esteem as evaluators since they did not see themselves as content experts, but rather as providing complementary expertise. In addition, the building administrators we talked to recognized that they needed help to do four to five required observations, so they welcomed the help of the peer evaluators. MCESA and its partner districts also responded to the concerns of building administrators that surfaced during the initial implementation. They increased communication about the program, emphasized the importance of building relationships with school administrators, retrained or released peer evaluators who had trouble with relationships with teachers or administrators, and changed scheduling practices. Field specialists provided additional support to building administrators, including providing training on the rubrics and aligned elements of the building leader rubrics, and did co-observations to help make administrators more comfortable with the more rigorous rubric and their competence as evaluators. They also responded to complaints and mediated conversations about teachers' ratings between building administrators and peer evaluators. As building administrators experienced these program improvements, and saw some of the positive impacts, opposition declined substantially.

The content specialization of the peer evaluators, the provision of coaching as well as feedback, and peer evaluators' involvement in providing in-school professional development have likely reduced teacher unease with peer evaluation. Content specialization provides credibility as well as the opportunity to get feedback from someone who has a deeper understanding of the subject taught than a building administrator might. The additional coaching and professional development provides teachers with resources to help them improve in areas the ratings suggest need work, and encourages teachers to see the peer evaluator as a supporter as well as an evaluator. Another factor that could promote teacher acceptance is the peer evaluators' employment outside of the district or school. This could make teachers less concerned that they would lose face with peers in their school or district if they received a less than perfect rating. The potential downside, teachers' fears that an evaluator from outside will not understand the school context, has been addressed by the additional effort the peer evaluators have put into becoming familiar with schools and participating in school professional development, professional learning communities, and leadership teams.

---

[3] It should be noted that due to opposition from its local teacher organization, one district dropped out of the TIF 3 grant, even though the teacher organization represented a minority of the district's teachers. However, peer evaluation is not the primary reason for the opposition.

Balancing Evaluation and Coaching

From an initial emphasis on rating and providing feedback, Maricopa County has moved peer evaluation toward coaching and providing teachers with professional development opportunities. This appears to have increased teacher support and also made the job more rewarding to the peer evaluators. However, it remains to be seen if over time peer evaluators become more lenient in their ratings, due to having invested more effort in their assigned teachers and wanting to see the effort pay off in improved ratings. Some features built into the Maricopa system are likely to discourage leniency. First, peer evaluators receive extensive and ongoing training on applying the observation rubric accurately and consistently. They are also evaluated on their use of evidence and have to pass yearly calibration tests. Second, the rigor of the selection process and the training emphasize the importance of accurate evaluation and honest feedback as well as providing a preview of the job demands. Some who are not comfortable with accurate rating are likely to have self-selected out, or been screened out during the selection process. Third, the degree of interaction among the peer evaluators and the emphasis on improving teaching seems to have created a culture among the peer evaluators that values accuracy as well as assistance. As mentioned above, the practice ratings provided to teachers by the combination of peers and administrators showed a substantial proportion rated less than effective.

Ensuring Consistency of Ratings and Feedback Between Peer Evaluators and Supervisors

As argued above, peer evaluation will lack credibility and confuse teachers if peer evaluators and building administrators apply different standards when rating teachers or recommend different changes in practice. Maricopa County's efforts to avoid this have included training and calibration testing of both peer evaluators and building administrators, having peer evaluators and field specialists do co-observations with administrators, and having peer evaluators share their ratings with building administrators after each observation. Peer evaluators make themselves available to discuss their observations and ratings with building administrators. In addition, similar elements related to teacher evaluation are part of both groups' performance evaluation rubrics.

Conclusion

Based on what we have learned from the Maricopa County program, we believe that other TIF grantees and school districts should consider peer evaluation as a potential contribution to effective educator evaluation. As practiced in the Maricopa County TIF districts, peer evaluation has the potential to support more rigorous performance evaluation as well as improve teaching practice by:

• Providing teachers with content-specific feedback and coaching to help them improve their performance

• Allowing for more observations, which in turn leads to a more reliable assessment of teachers' practice, gives teachers more feedback, and provides multiple opportunities for teachers to apply the feedback and coaching suggestions

• Helping educate building administrators about how to observe teachers and apply the rubrics and providing them with a yardstick against which to measure their own ratings of teaching

• Improving the motivation of building administrators to make accurate evaluation ratings by building their efficacy as observers, confirming their evidence gathering, and sharing the burden of having to deliver results to teachers with performance problems.

It is yet to be seen whether peer evaluation will substantially improve instructional practice in Maricopa County's participating districts. The TIF program evaluation, when completed, should provide some evidence. Anecdotal evidence collected by MCESA staff suggests that some teachers improved their practice as a result of feedback and coaching by peer evaluators, that some teachers felt the process re-invigorated their teaching, and that peer evaluation coupled with a more rigorous evaluation process may have encouraged marginal teachers to leave. However, peer evaluation appears to have contributed to a less skewed distribution of teacher evaluation ratings, in contrast to the tendency seen in many states and districts to have the vast majority of teachers rated effective or highly effective.

Many of the barriers to using peer evaluation can be surmounted with careful planning and program design. To make peer evaluation work, other TIF grantees and school districts interested in using peer evaluation should consider:

- Developing a hiring process that selects people with content, coaching, and relationship-building skills

- Providing ongoing rater calibration training to maintain inter-rater agreement and consistency of feedback

- Training peer evaluators in coaching and relationship building skills as well as in evaluating accurately

- Having peer evaluators go beyond providing ratings and feedback to do coaching and professional development with the teachers they evaluate

- Communicating about the program and building relationships with teachers and school leaders

- Establishing multiple sources of feedback about how peer evaluation is being implemented and received at the school level.

An innovation like peer evaluation also needs dedicated support. In Maricopa County, MCESA's field specialists played a key role in making peer evaluation work. MCESA created the job of field specialist to be the TIF program's primary contact person with each district. The field specialist is "on call" and able to step into issues quickly. They perform a wide range of tasks, including communication liaison between the peer evaluator, building administrator, and district central office staff and helping building administrators understand the evaluation instruments, coaching, and observing. The field specialists also facilitated resolution of issues and differences involving evaluation scores and helped develop shared understandings. Field specialists conducted co-observations with building administrators and helped develop job-embedded professional development plans. Also, the field specialists had informal and private conversations with the building administrators about problems teachers may have with performance, peer review, or the evaluation system. Finally, the field specialists participated in the recruitment and selection of the peer evaluators, communicated with school boards, parents and the press, and aided districts in pay-for-performance program design. As channels of communication to and from districts, they were a key factor in resolving the early implementation issues peer evaluation faced. This kind of dedicated support would be a valuable resource when implementing peer evaluation, as well as other parts of a TIF program.

References

Chambers, J. G., Lam, I., & Mahitivanichcha, K. (2008). *Examining context and challenges in measuring investment in professional development: A case study of six school districts in the southwest region. issues & answers. REL 2008-no. 037.* San Antonio, TX: Regional Educational Laboratory Southwest.

Meyer, H. H, Kay, E., & French, J. R. P. Jr. (1965). Split Roles in Performance Evaluation. *Harvard Business Review*, 43, 123–129.

Milanowski, A. T., (2005). Split Roles in Performance Evaluation - A Field Study Involving New Teachers. *Journal of Personnel Evaluation in Education*,18:3, 153–169.

Miles, K. H., Odden, A., Fermanich, M., & Archibald, S. (2004). Inside the black box school district spending on professional development in education: Lessons from five urban districts. *Journal of Education Finance*, 30(1), 1–26.

Popham, W. J. (1988). Judgment-Based Teacher Evaluation. In S. J. Stanley & W. J. Popham (eds.), *Teacher Evaluation: Six Prescriptions for Success*. (pp. 56–71). Alexandria, Virginia: Association for Supervision and Curriculum Development.

Project on the Next Generation of Teachers. (n.d.) *A user's guide to peer assistance and review*. Harvard Graduate School of Education, **http://www.gse.harvard.edu/~ngt/par/**

Odden, A., Archibald, S., Fermanich, M., & Gallagher, A. H. (2002). A cost framework for professional development. *Journal of Education Finance*, 28(1), 51–74.